DOCUMENT RESUME

ED 093 918 TM 003 735

AUTHOR Besel, Ronald; Okada, Masahito
TITLE The Development of Domain-Referenced Tests for an Objectives-Based Reading Program.
PUB DATE [Apr 74]
NOTE 8p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Criteria; *Reading Tests; *Test Construction
IDENTIFIERS *Domain Referenced Tests

ABSTRACT

Criteria for the selection of item forms, content domains, and sampling procedures for program specific, domain-referenced tests are developed. The primary purpose of these tests is to estimate the extent to which individual pupils have attained or retained the intended learning outcomes of a particular segment of instruction. Tests developed for the tryout of the SWRL Reading Program illustrate the application of the criteria. A variety of critical reading skills is assessed. The use and potential value of facet designed tests for assessing word recognition and novel word decoding is described. Error type scores provide potentially valuable information on which to base prescriptions of supplementary instruction. (Author)

ED 093918

Session #26.02.

THE DEVELOPMENT OF DOMAIN-REFERENCED TESTS FOR AN OBJECTIVES-BASED READING PROGRAM

Ronald Besel and Masahito Okada

SWRL Educational Research and Development

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 003 .35

ABSTRACT

Criteria for the selection of item forms, content domains and sampling procedures for program specific, domain-referenced tests are developed. The primary purpose of these tests is to estimate the extent to which individual pupils have attained or retained the intended learning outcomes of a particular segment of instruction. Tests developed for the tryout of the SWRL Reading Program illustrate the application of the criteria. A variety of critical reading skills are assessed. The use and potential value of facet designed tests for assessing word recognition and novel word decoding are described. Error type scores provide potentially valuable information upon which to base prescriptions of supplementary instruction.

THE DEVELOPMENT OF DOMAIN-REFERENCED TESTS FOR AN OBJECTIVES-BASED READING PROGRAM

Tests for assessing a variety of important reading skills are a critical component of an objectives-based reading program. These tests have as their primary purpose the estimation of the extent to which individual pupils have attained or retained the intended learning outcomes of a particular segment of instruction. This information provides a basis upon which teachers can prescribe needed supplementary instruction and adjust the pace of instruction. The following criteria have proven useful in the selection of item forms, content domains and sampling procedures.

1. If time used for assessment purposes is to be justified, then it is essential that the data generated by testing effect instructional decision making. Outcomes selected for tests should be judged to be critical to the future success of students on reading tasks. The testing of outcomes for which no supplementary instructional materials or remedial procedures are available should be avoided.

2. In order to minimize the confounding effects which may result due to the use of novel item forms, test items should relate directly to instructional materials and procedures used in the program. Item forms should not be used unless an identical or very similar task has been previously used in instruction.

3. Since supplementary instruction is one of the primary reasons underlying regular assessment practices, it is essential that the content domain for assessment be of manageable proportions in respect to such instruction. One way to achieve this is to separately assess skill achievement within the content domain of the most recent instruction from previously taught content.

4. Although most instruction may be group based, to permit individualized instruction for those students requiring supplementary instruction outcome scores should provide valid information on individual pupils. Test length should be adequate to provide reliable pupil scores for each outcome.

5. Content validity is desirable. Explicit statements defining the content domains should be formulated and a defensible sampling procedure should be employed to assure that the sample of test items is representative of the content domain for a unit of instruction.

6. Whenever possible, the distractors for multiple-choice items should be systematically selected to represent meaningful error types. This not only reduces the possibility that irrelevant factors influence what the test measures but allows for diagnostic scoring of the test. Error type scores can then be used in the selection of supplementary instruction.

The tests used in the initial tryout of the SWRL Reading Program are illustrative of the application of these criteria. The systematic

selection and sequencing of phonic skills and vocabulary content which was employed in developing the SWRL Reading Program provides a rational basis for specifying the content domains for domain-referenced assessment. The first task in developing the program was the selection of a lexicon appropriate for children at the K-3 level. The resulting lexicon consisted of approximately 9000 words which were believed to be in the recognition, if not active, vocabulary of K-3 children. Extensive research on English orthography, including the study of spelling-to-sound correspondences for the 6000 one and two syllable words from the lexicon, resulted in the sequencing of vocabulary for K-3 reading. (Cronnell, 1973.) Words consistent with this sequence were then selected for the stories to be written for each level of the program. At a later date additional words were selected for such program components as phonics instruction and supplementary instructional activities called Practice Exercises.

Instructional outcomes selected for testing include: letter names; word recognition; novel word decoding; word meaning; sentence comprehension, and paragraph comprehension. Instruction and testing on letter names is confined to the initial instructional block. To make tests of manageable length, assessment of paragraph comprehension is deferred until Block 5 by which time the students have been given practice in the test task. Novel word decoding is likewise deferred until sufficient phonics instruction is completed. The testing of novel word decoding is discontinued at the end of Block 4 since the distinction between novel words and words the pupil is likely to recognize becomes increasingly difficult to make.

The task used to assess word recognition within Blocks 1-4 has been demonstrated to pose little difficulty for the child beginning to read; the teacher reads a word and a sentence containing the word and the student must select the correct word from a set of four options, all of which are previously taught words. Performance is assessed for two content domains; (a) words introduced in the storybooks of the current unit and (b) words introduced in the immediately preceding unit. Supplementary instructional materials are available for each of these domains. The novel word decoding task appears to be similar but the correct option is a word from a defined lexicon of words in the passive vocabulary of pupils age 6-7 which has not been used in any previous instructional materials. The distractors were picked from the same lexicon with storybook words deleted. Thus, most pupils will be required to sound out each option and compare it with the stimulus.

In Blocks 5-8 a combined word recognition--word meaning task is used to assess vocabulary skills for three domains of content: (a) words introduced in the storybooks of the current unit, (b) words introduced in the immediately preceding unit, and (c) "vocabulary extension" words which were used in workbook activities. The task requires the student to select the word which best completes a short phrase or sentence. Use of this task, rather than the Blocks 1-4 recognition tasks, reflects the increasing instructional emphasis on word meaning and a declining emphasis on oral word decoding in the upper levels of the Program. Since supplementary instruction can deal with both outcomes simultaneously, little is lost in the use of a test task requiring both skills.

The domain of words used in vocabulary extension activities varies systematically as instruction proceeds. In Block 5 it consisted of words instructionally paired with a word introduced in a story which has a high graphemic and/or phonemic similarity. Families of words having a common phoneme or word part constituted the Blocks 6 & 7 domains. Block 8 used words having the same root. Delimiting the content domains in this way made it possible to design supplementary instruction which covered the same domain as the domain-referenced tests.

Sentence comprehension was tested in Blocks 1-4 using modified cloze items consisting of sentences which were representative of the syntactical structures previously used in Storybooks. Paragraph comprehension was considered to be a higher order skill and requiring more test-taking skills on the part of the student to permit valid assessment. Therefore assessment of paragraph comprehension, using true-false statement referenced to a short story, was deferred until Block 5. Multiple choice questions were introduced in paragraph comprehension assessment in Block 8.

The selection of distractors for the Blocks 1-4 word recognition and novel word decoding tasks followed a specified facet design. Each facet represents a meaningful error type. The content of a test item is assumed to have two aspects: the stimulus and the response options. A facet is defined to be a characteristic on which the stimulus and a option can be evaluated and compared. Any one syllable word can be conceptualized in terms of three facets; an initial consonant sound, a medial vowel sound and a final consonant sound. A facet design is a specification of the desired patterns of similarity between distractors and stimuli accompanied with substitution rules for when a desired distractor is non-existent.

The facet design chosen specified the selection of one distractor of each of three high similarity classes: (a) an initial consonant distractor (differs from the stimulus only with respect to initial consonant sound) (b) a medial vowel distractor and (c) a final consonant distractor. Most of the Block 1-4 words are one syllable so the facet design could be employed for most of the test items. Analysis of kindergarten student responses to similar items within the SWRL Beginning Reading Program indicated that there may have been differential learning with respect to the three error types (Besel, 1972a). The frequency with which students picked initial consonant distractors declined rapidly as the student progressed through the program while the medial vowel and final consonant error rates remained relatively constant over time. More recent placement test data from students at the completion of first grade is consistent with the trend in kindergarten error rates. Most of the high attraction distractors were from the medial vowel class. The only initial consonant distractors which were highly attractive were those which had the same initial letter as the stimulus and differed only in subsequent consonants of an initial consonant cluster or digraph.

Analysis of kindergarten posttest data employing a cluster analysis of error-type scores indicated that there were subgroups of students exhibiting error patterns which could not be identified solely on the basis of outcome scores (Besel, 1972b). Error type scores thus appear to provide potentially valluable information for prescribing supplementary instruction.

The Blocks 1-4 tests were designed to provide a means for confirming these earlier results in a context where the additional information obtained from error type scores can be evaluated and exploited.

References

Besel, R., 1972a. An initial facet analysis of the FYCSP word attack test. Technical Note No. TN-5-72-46, August 18, 1972. Southwest Regional Laboratory, Los Alamitos, California.

Besel, R., 1972b. Cluster analysis of error score data. Technical Note No. TN-3-72-40, December 1, 1972. Southwest Regional Laboratory, Los Alamitos, California.

Cronnell, Bruce. Designing a Reading Program Based on Research Findings in Orthography. Elementary English, January, 1973, pp 27-34.